xLiMe
FP7-ICT-2013-10
ICT-2013.4.1
Contract no.: 611346
www.xlime.eu

Deliverable D4.3.2

Final Semantic Graph Construction Prototype

Editor:	Aljaž Košmerlj, JSI
Author(s):	Aljaž Košmerlj, JSI, Blaž Fortuna, JSI;
Deliverable Nature:	Report (R)
Dissemination Level:	Public (PU)
Contractual Delivery Date:	M30 – 30 April 2016
Actual Delivery Date:	M30 – 30 April 2016
Suggested Readers:	All project partners
Version:	1.0
Keywords:	text mining; natural language processing; semantic graph extraction

Disclaimer

This document contains material, which is the copyright of certain xLiMe consortium parties, and may not be reproduced or copied without permission.

All xLiMe consortium parties have agreed to full publication of this document.

The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the xLiMe consortium as a whole, nor a certain party of the xLiMeconsortium warrant that the information contained in this document is capable of use, or that use of the information is free from risk, and accept no liability for loss or damage suffered by any person using this information.

Full Project Title:	xLiMe– crossLingual crossMedia knowledge extraction
Short Project Title:	xLiMe
Number and Title of Work package:	WP4 Cross-media Semantic Integration
Document Title:	D4.3.2 - Final Semantic Graph Construction Prototype
Editor:	Aljaz Kosmerlj, JSI
Work package Leader:	Aditya Mogadala, KIT

Executive Summary

This deliverable presents the final prototype for semantic graph construction developed within the xLiMe project. As opposed to the earlier work (presented in deliverable D4.3.1), where the focus was on unsupervised induction of semantic graphs, we formulate the problem as a supervised task of filling nodes of given template semantic graphs for individual event types. An example of this would be to find for event type 'soccer game' the score, name of the winning team, name of the losing team, names of players who scored goals etc.

Our approach to this task is to build classifiers for filling individual slots which use features computed through aggregation over event clusters obtained from the Event Registry system. As far as we know we are the first to employ this approach as other state-of-the-art methods that we are familiar with use only features from single articles. Our method can also work in cross-lingual scenario by using all articles from cross-lingual Event Registry clusters. This is achieved by using a statistical method for semantic similarity computation implemented in the Xling system to project language dependant bag-of-words features of examples from different languages into a common space. This allows them to be processed by the same classifier.

For evaluation purposes we focus on two event types: earthquakes and company acquisitions. We build labelled datasets of Event Registry events using external data sources. Results obtained on these datasets through cross-validation show that our approach can achieve results comparable to state-of-the-art slot filling methods in a mono-lingual setting. In a cross-lingual setting the scores drop a little, but are still reasonably high. In conclusion we identify several options for future improvements of this approach, most notable one being the integration of this methodology into Event Registry and enabling users to efficiently build custom slot fillers using active learning.

Table of Contents

List of Figures

List of Tables

Abbreviations

A list of abbreviations is strongly recommended

API	Application Programming Interface
EMSC	European-Mediterranean Seismological Centre
ER	Event Registry
tfidf	term frequency – inverse document frequency
SVM	Support Vector Machine
CV	Cross Validation

1 Introduction

This deliverable presents the final version of the prototype for extracting structured information from unstructured data developed in the scope of xLiMe project [1]. The target structured data representation for extraction is a semantic graph – a knowledge representation which consists of a set of concepts (vertices) and relations between them (edges). Such a structure can easily be parsed and processed automatically by machines and is highly useful for data sharing and interchange across applications, systems and organizations as well as various data processing and storage tasks.

In preceding work [2] we focused on unsupervised construction of semantic graphs by aggregating semantic frames extracted from news articles describing the same event. We used a Wordnet based semantic similarity measure to match frames with equivalent meanings and through redundancy offered by multiple articles identified and pruned parts of the graph with low support in the data. The major drawback of that approach was the lack of cross-lingual support and the necessity for labour-intensive manual evaluation for the experiments.

To avoid problems with manual evaluation, we reformulated the problem into a supervised form. The assumption is that the user identifies the target event type (e.g. bombing attack, product recall, soccer game etc.) and its template semantic graph (i.e. typical roles concepts and entities play in the events of this type and the relations between them). The problem we are solving is to extract entities and concepts that fill individual event type roles from cross-lingual clusters of news articles about the same real-world event.

Please note that throughout this document we use the term ‘event’ for both the article cluster as well as the real-world happening it represents. Where a distinction might be important we point it out explicitly.

1.1 Relation to Other Work Packages and Deliverables

This deliverable uses semantic annotations from tools developed in work packages WP3 and WP4 in order to extract a semantic representation of events from unstructured data (i.e. news article clusters).The extracted output is suitable for machine processing and is intended for consumption in the analytics tools developed in work package WP5. The relations are summarized in Table 1.

Table 1: Relation to other work packages

Component (Core Functionality)	Receives input from WP	Provides input to WP/D
Structured event information extractor	Semantic annotation and integration tools developed in work packages **WP3** and **WP4** described in their respective deliverables.	Structured event information to be used in analytics tools developed in work package **WP5**.

1.2 Overview of the Document

The remaining document is structured as follows. In section 2 we present the developed method for extraction of structured event information. We specify the problem in greater detail and highlight the innovations contributed as well as the key performance indicators (KPI’s) used to measure them. Subsection 2.2 contains a detailed technical description of the methodology used and in subsections 2.1 and 2.3 we describe the datasets used in our experiments and the results obtained by our evaluation. Section 3 contains several ideas for future work and finally we conclude in section 4.

2 Extraction of Structured Event Information

Online news represents a vast data source that is unfortunately hard to process automatically due to the unstructured nature of its text. The Event Registry system makes great strides in this direction by clustering articles by content and extracting some general structured information from them as well as categorizing them into topic categories using the DMOZ taxonomy. However, obtaining event type specific information, such as the score of a football game, the number of casualties of a bombing attack or the product recalled in a product recall, is still not automatically available. Such information would enable a new level of insight into real-world trends and would be invaluable in the areas such as finance and social policy. The problem can be stated as follows:

***Assuming we know a given event (i.e. article cluster) belongs to some event type, we would like to automatically extract event type specific information and return it in structured form*.**

The assumption that we know the event type of a given event is highly non-trivial (in a general automated setting). The approach we describe relies entirely on our access to categorized events for learning the extractor classifiers and although we could apply them to any event, such a blind application is unlikely to be reliable. For our analysis we semi-automatically construct datasets from external event data using Event Registry API in order to obtain data with labelled slots. In general, this problem can be solved using the *Categorizer* system described in Deliverable D5.3.2 [4] which allows users to semi-automatically create custom event type categorization classifiers using active learning.

In this formulation the problem is essentially a slot-filling or knowledge base population task which is a well-known problem. The Text Analysis Conference[1] (TAC) organized yearly by U.S. National Institute of Standards and Technology (NIST) has hosted a knowledge base population competition for the last several years [5][6], with even a special track for Event Argument Extraction and Linking in 2016[2]. Top-ranking submissions in recent years use deep learning [7][8] and distant supervision [9] to achieve their results. All the competitions use a single-article-extraction setting and to the best of our knowledge there are no other approaches which extract structured information from events (i.e. article clusters). There are also no gold standard datasets for event extraction that we know of, so we prepared our own for the purpose of this deliverable.

The problem definition is summarized in Table 2 along with key performance indicators used to measure success.

Table 2: List of problems and relevant key performance indicators (KPI's)

Problem Definition	Objective Target (Evaluation Measure)	Measureable Progress
Structured event information extraction from (cross-lingual) clusters of news articles.	Standard classification measures: precision, recall, F1.	In a mono-lingual setting we achieve results comparable with state-of-the-art methods. In a cross-lingual setting the performance drops, but is still reasonable with respect to state-of-the-art.

2.1 Datasets

In our experiments we focus on two event types: **earthquakes** and **company acquisitions**. These were chosen because they are commonly reported in the media and have a typical structure that can be described as a semantic graph. We present the dataset for each event type individually.

[1] http://www.nist.gov/tac/

[2] http://www.nist.gov/tac/2016/KBP/Event/index.html

2.1.1 Earthquake

Earthquakes are natural disasters common all around the world and range in intensity from minor annoyances to major natural and humanitarian catastrophes. Earthquakes of medium and high intensity levels are widely covered in the media all over the world, not just locally where they occur. In order to build a supervised dataset we used the European-Mediterranean Seismological Centre (EMSC) online search interface[3]to obtain a list of all earthquakes felt anywhere in the world in the years 2014 and 2015(specifically from 1.1.2014 to 31.12.2015). The list contained information about 3190 earthquakes, but included also earthquake aftershocks, which are not typically reported about specifically, as separate entries.

Each earthquake was described with the following features: UTC date-time, latitude and longitude, depth, depth type, magnitude, magnitude type, region name, last update, eqid (i.e. earthquake id). Of these only date-time, location (i.e. region name), magnitude and depth are typically reported in news media. Date-time and depth proved challenging to accurately automatically annotate in the article text because date-time is commonly implied using relative terms (i.e. ‘yesterday’, ‘last Tuesday’) and reported depth values commonly did not match the values obtained from the EMSC, likely due to different sources of seismic measurements. We focus on location and magnitude for our analysis and the (small) template graph is shown in Figure 1. Note that the location is a concept (denoted in the figure as a round node) and magnitude is a literal (numeric) value (denoted in the figure as a rectangular node).

In order to obtain Event Registry (ER) events for the earthquakes listed in the EMSC dataset we first used the ER Python API to obtain Wikipedia location URI’s from region names reported in the dataset. We then queried ER for all events in the window of +/- one day around the date from the ESMC using the location concept and the earthquake concept (http://en.wikipedia.org/wiki/Earthquake) as query conditions. We manually inspected top 3 results for each query (not all queries returned more than 3 events and some returned none at all) to confirm they are indeed a match for the earthquake. We thus obtained 96 events with total of 4141 English, 2350 Spanish and 794 German news articles, averaging roughly 76 articles per event.

For each article we identified a set of slot candidates which consisted of all annotations obtained using the text annotations from WP3 [3] and all numerical phrases. A numerical phrase is any non-whitespace delimited string which contains a number (e.g. ‘19’, ‘23:46’, ‘3.1-magnitude’). If such a word was followed by a number word such as ‘thousand’ or ‘billion’ we included that word into the phrase as well. Positive location slot fillers were trivial to identify since we had beforehand matched the locations to Wikipedia URIs which are used for the annotations. Magnitudes were identified using regular expression matching with numerical phrases and then by comparing values to those from the EMSC dataset (with an error margin of 0.1 allowed).Numbers of slot candidates per language and slot are listed in Table 3.

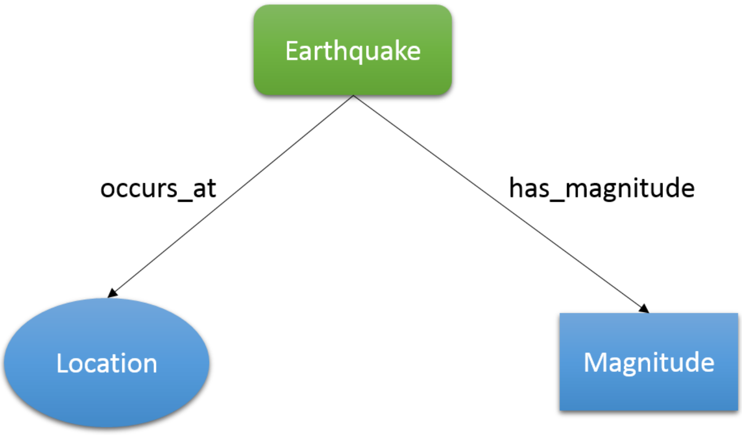

Figure 1: Earthquake template graph

[3] http://www.emsc-csem.org/Earthquake/?filter=yes

Table 3: Numbers of slot candidates per language for earthquake events

language	*none*	*location*	*magnitude*
eng	23002	401	102
spa	6994	95	52
ger	1404	20	3

2.1.2 Company Acquisition

An acquisition is a corporate action in which a company buys most, if not all, of the target company's ownership stakes in order to assume control of the target company. They commonly represent major economic events with millions or even billions being paid for ownership and are widely reported, especially when high profile companies (e.g. Apple, Pfizer, Nestle) are involved. We obtained a dataset of top 5000 acquisitions (sorted by amount paid) in years 2014 and 2015 from Bloomberg L.P., a leading financial software, data, and media provider, and used it to build a labelled dataset.

Each acquisition was described with an announcement date, target company name, acquirer company name and total value paid along with some metadata which is not interesting for our analysis. A template semantic graph for the acquisition event is presented in Figure 2. We obtained an events dataset by querying ER similarly as before, but linking the acquisitions to events proved much more challenging than with earthquakes. We queried ER for Wikipedia URIs of target and acquirer companies. The coverage of companies by Wikipedia is much worse in comparison to Geographical names and company names are more likely to be similar to other concept names. Also it is common for companies to change names after they have been acquired. The Bloomberg dataset contained updated names which are typically not the ones that were reported in the media when the acquisition occurred. Finally, acquisitions are less atomic events with several stages (e.g. announcement, negotiation, and transfer of ownership) spread across a longer time. In order to maximize the chances of obtaining relevant events, we queried ER for all events from the acquisition announcement on, with Target Company and acquirer company URIs as well as the acquisition concept (http://en.wikipedia.org/wiki/Mergers_and_acquisitions) as conditions. As with earthquakes we took into consideration only top 3 results from each query and manually cleaned out the bad events. In the end we obtained 71 events with total of 4279 English, 174 Spanish and 327 German news articles, averaging roughly 67 articles per event.

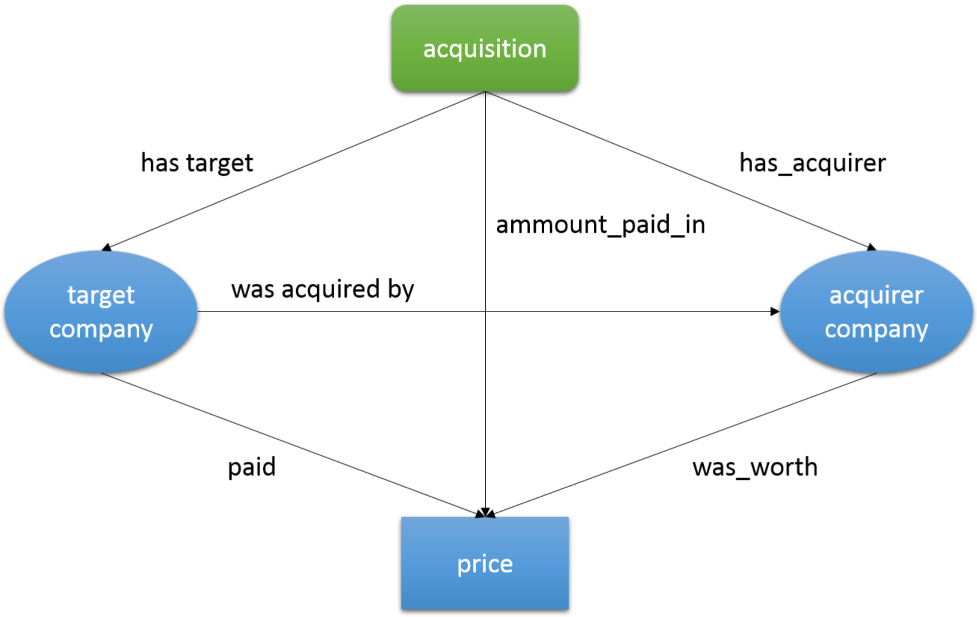

Figure 2: Acquisition template graph

Slot candidates were identified in a similar manner as with the earthquakes dataset. Since reported values for price paid in the acquisition are commonly rounded or a little erroneous the matching criterion for numerical phrase values with price was more lax in comparison to earthquake magnitudes and the reported values could differ from the true value for 15 %. The final numbers of slot candidates per language and slot are listed in the Table 4.

Table 4: Numbers of slot candidates per language for acquisition events

language	*none*	*acquirer*	*target*	*value*
eng	38971	39	37	48
spa	918	4	2	0
ger	1000	7	6	0

2.2 Methodology

The problem of filling slots of template graphs, as we have stated it, is a supervised classification task. For each template graph (i.e. each event type) we build a classifier that determines if a slot candidate fills any slot (node). As explained in the previous section, slot candidates are all strings in the articles annotated with the text annotation tool from WP3 as well as all numerical phrases. We tested two settings of our experiments: mono-lingual and cross-lingual. The mono-lingual setting shows the effect of using data redundancy offered by clusters of articles in comparison to state-of-the-art slot filling methods work on single articles and also serves as a baseline for the cross-lingual setting which leverages all the articles in the clusters regardless of the language. The two settings differ only in the computation of slot candidate features, which is described in the remainder of this section along with learning methodology.

2.2.1 Mono-lingual setting

In this setting only articles from one language are taken into account at a time. For each slot candidate the following set of features is computed:

- *Portion of containing articles*–The percentage of articles in the event cluster that contain the slot candidate. The feature is discredited into 11 regions which roughly correspond to value distribution: 0-1, 1-2, 2-3, 3-5, 5-7, 7-10, 10-15, 15-20, 20-30, 30-50, 50-100, resulting in 11 binary features.
- *Number of containing articles* – The nominal number of articles in the event cluster that contain the slot candidate. The feature is also discredited into 11 regions: 1-2, 2-3, 3-5, 5-7, 7-10, 10-15, 15-20, 20-30, 30-50, 50-100, 100-, resulting in 11 binary features.
- *Position distribution type* – We computed the distribution of slot candidate mention positions over article text and aggregated over the entire event cluster. The distribution is computed as a quadruple (p_1,p_2,p_3,p_4) where p_i is the portion of mentions occurring in the i-th quarter of the articles text. The distribution is compared to 14 prototype distributions and the closest one (using Euclidean measure to compare distributions) is determined. 14 binary features are constructed, one for each prototype distribution, and the value of the closest one is set to 1 and the others to 0. Prototype distributions are listed in Table 5.
- *Bag-of-words features* – We collected 5 words (i.e. whitespace-delimited substrings) from the article text before and after each slot candidate mention over all the events into two multi-sets (bags): bow_left and bow_right. Each multi-set is the used to produce tfidf weighted features. The number of produced features varies depending on the dataset but was typically in the [5000, 30000] range. Due to their nature, these features are very sparse.

Table 5: Prototype distributions

distribution name	p_1	p_2	p_3	p_4
uniform	0.25	0.25	0.25	0.25
1st quarter	1	0	0	0
2nd quarter	0	1	0	0
3rd quarter	0	0	1	0
4th quarter	0	0	0	1
1st half	0.5	0.5	0	0
2nd half	0	0	0.5	0.5
center	0	0.5	0.5	0
not 1st quarter	0	0.3333	0.3333	0.3333
not 4th quarter	0.3333	0.3333	0.3333	0
linear drop	0.4375	0.3125	0.1875	0.0625
linear rise	0.0625	0.1875	0.3125	0.4375
exponential drop	0.675	0.225	0.075	0.025
exponential rise	0.025	0.075	0.225	0.675

2.2.2 Cross-lingual setting

In the cross-lingual setting articles of all languages are taken into account. Looking at the features from the mono-lingual setting it is clear that all but the bag-of-words features are completely language agnostic. Technically we could even use an identical approach and just collect words from all languages into the same bags of preceding and succeeding words; however that would vastly increase the feature space as well as the amount of necessary learning examples. A better approach is to take semantics into account to avoid modelling the same context meaning in each language.

To achieve this we use Xling[4][10] a cross-lingual semantic similarity measure based on latent Semantic Indexing and Canonical Correlation Analysis. The details of the measure far exceed the scope of this deliverable. It is sufficient to understand that it is capable of projecting text from different languages (top 100 languages by size of native-language Wikipedia corpus) into a common space (a statistical "middle language" of sorts). We use this approach to project preceding and succeeding bag-of-words instances for all slot candidates into a comparable form, replacing them with 500 real-valued features (normalized to [-1,1]). Since all examples now share the same feature space, we can use the entire dataset for learning.

2.2.3 Learning

We used a Support Vector Machine (SVM) with a linear kernel as classifier, l2 regularization and class weighting inversely proportional to class frequencies. We built one classifier per template graph using one-vs-all approach for cases with more than one slot. The classifiers were learned using stochastic gradient descent (20 learning iterations). The entire experiment was implemented in Python using scikit-learn module, except for the bag-of-words' expansion to tfidf weighted features, which was computed in Javascript using the Qminer[5] platform for convenience.

2.3 Evaluation

For evaluation, we performed 5-fold cross validation (CV) for each event type and language separately in the mono-lingual setting and for each event type separately in the cross-lingual setting. All folds were

[4] http://xling.ijs.si/

[5] http://qminer.ijs.si/

stratified by class. Precision, recall and F1, all standard classification success measures, were computed. The results obtained are presented in discussed in the remainder of this section.

Mono-lingual Setting

Results obtained for the earthquake and acquisitions are presented in Table 6 and Table 7 respectively.

Table 6: Results for earthquake events in the mono-lingual setting

	eng			spa			Deu		
	prec	recall	F1	prec	recall	F1	prec	recall	F1
none	0.991	0.961	0.976	0.991	0.971	0.981	0.985	0.981	0.983
magnitude	0.299	0.571	0.371	0.274	0.537	0.356	0.240	0.100	0.124
location	0.300	0.725	0.411	0.407	0.673	0.490	-	-	-

Table 7: Results for acquisition events in the mono-lingual setting

	eng			spa			deu		
	prec	recall	F1	prec	recall	F1	prec	recall	F1
none	0.998	0.985	0.992	0.993	0.992	0.993	0.990	0.988	0.989
acquirer	0.133	0.411	0.197	0.000	0.000	0.000	0.000	0.000	0.000
target	0.204	0.621	0.291	-	-	-	0.300	0.300	0.267
value	0.096	0.127	0.094	-	-	-	-	-	-

Unfortunately, there were not able to compute evaluations for all slots across all languages as there were not enough slot examples in all languages (see Table 3 and Table 4). Results for the '*none*' slot are the results of the classifier for the negative examples (i.e. the slot candidates that do not fill any slot). We left them in the table for completeness, but they are not very informative.

Otherwise the results we obtained are promising. The reported F1 scores of state-of-the-art methods for slot filling on single articles is 0.367 [6] (with median results from competitions closer to 0.2) and those are results obtained on larger learning datasets and with a lot of tuning of classifiers, whereas we obtained ours with little to no classifier tuning (the settings we used are mostly default settings for the SVM classifier in the scikit-learn module).

By far the worst result is obtained on the *value* slot of the acquisition event type. Manual inspection has shown that the vocabulary used around mentions of values in acquisitions indeed varies much more as for example with earthquake magnitudes, which were also annotated as numerical phrases. We believe the biggest reason for the poor performance however is the relatively high error rate of the matching of true values from the Bloomberg dataset to the text annotations. Since reported values often significantly deviate from the true value we had to allow for larger differences when looking for matches who resulted in more spurious matches and more noise.

Cross-lingual Setting

Moving to the cross-lingual setting the results are no longer grouped by language as the entire dataset for each event type is now used at once. Results for earthquakes and acquisitions in this setting are reported in Table 8 and Table 9 respectively.

Table 8: Results for earthquake events in the cross-lingual setting

	prec	recall	F1
none	0.992	0.947	0.968
magnitude	0.202	0.563	0.286
location	0.248	0.727	0.367

Table 9: Results for acquisition events in the cross-lingual setting

	prec	recall	F1
none	0.999	0.983	0.991
acquirer	0.139	0.574	0.219
target	0.126	0.440	0.183
value	0.041	0.164	0.061

Maximum scores in both event types drop a little but are still decent. The drops in scores are mostly due to drops in precision, whereas on the other hand recalls mostly increase or stay comparably high (with the exception of the target slot in acquisitions). Since recalls increase in comparison to the values obtained on the English dataset and since the cross-lingual setting has more positive examples it seems that sharing the feature space with English has enabled the classifiers to capture more values in other languages and not the opposite (i.e. that the inclusion of data from other languages would hurt recall in English events).

3 Future Work

There are a lot of possible improvements we could make to our approach. The simplest one is tuning the classifiers used and adding additional features. The features used so far are all quite shallow and we could try adding semantic features such as types of annotations as well as value ranges and presence of units or currency symbols for numerical phrases. Such feature engineering could have a lot of impact especially if we have specific target event types in mind and insight into them (either on our own or through a domain expert).

A different option is to add a post-processing step where we determine which of the possible slot candidates returned by the slot filler classifiers most likely belong together in the template graph. Features such as relative position in the article text as well as semantic features using external knowledge bases could help significantly prune the result set and improve precision.

Finally, even though we improved evaluation from the early prototype, where it was limited due to the fact that it had to be entirely manual, it is still not at a satisfying level. We were somewhat disappointed with how hard it was to use external data sources to build a labelled dataset. The nature of the problem is such, that reported values in the articles are often shortened (company names), rounded (acquisition prices) or simply wrong as the real facts may not have been known at the time the article was written. We believe an answer to this problem lies in active learning methodology. We plan to extend the Event Registry Categorizer module, which can use active learning to learn event type classification, to be able to learn slot filling classifiers. Such a system could guide the user performing the annotations, ensuring best results for minimal amount of work.

4 Conclusion

We have presented the final prototype for semantic graph construction developed within the xLiMe project. In its scope we focused on filling individual slots of given template graphs for individual event types. This allowed us to formulate the problem as a supervised task which greatly simplified evaluation.

The method for extracting structured information (i.e. filling graph templates) we developed follows the standard approach of building classifiers for individual slots, but uses features aggregated over a cluster of articles about the same event whereas other methods work on features computed on single articles .Results in a mono-lingual setting show that this enables us to obtain state-of-the-art comparable results even with little datasets as the information redundancy offered by event clusters results in more reliable features. We have also adapted our approach to work in a cross-lingual setting by projecting language-dependent bag-of-words features into a common feature space using Xling, a statistical method for computation of cross-lingual semantic similarity of text. Results have shown that the adapted method works reasonably well, but there is still vast room for improvement.

Trends in the scientific community indicate that event extraction is gaining in popularity and interest, for example with specialized tracks for event extraction being organized at the Text Analysis Conference, and industry has been interested in this problem for a long time. Media analysis companies such as Bloomberg sell structured information about financial events for profit and this information is currently mostly extracted from news manually by people. All these circumstances show that this area of research will likely evolve a lot in the short to medium term. Results presented in this deliverable indicate that xLiMe partners hold resources (Event Registry, Xling etc.) that present promising avenues for development in this area and we plan to fully exploit this opportunity.

References

[1] http://www.xLiMe.org

[2] xLiMe deliverable D4.3.1 – Early Semantic Graph Construction Prototype

[3] xLiMe deliverable D3.3.2 –Final Prototype for Text Annotation

[4] xLiMe deliverable D5.3.2 –Final Analytics Prototype

[5] Surdeanu, Mihai. "Overview of the tac2013 knowledge base population evaluation: English slot filling and temporal slot filling." In *Proceedings of the Sixth Text Analysis Conference (TAC 2013)*. 2013.

[6] Surdeanu, Mihai, and Heng Ji. "Overview of the english slot filling track at the tac2014 knowledge base population evaluation." In *Proc. Text Analysis Conference (TAC2014)*. 2014.

[7] Angeli, Gabor, Sonal Gupta, Melvin Jose, Christopher D. Manning, Christopher Ré, Julie Tibshirani, Jean Y. Wu, Sen Wu, and Ce Zhang. "Stanford's 2014 slot filling systems." *TAC KBP* (2014).

[8] Niu, Feng, Ce Zhang, Christopher Ré, and Jude W. Shavlik. "DeepDive: Web-scale Knowledge-base Construction using Statistical Learning and Inference." *VLDS* 12 (2012): 25-28.

[9] Roth, Benjamin, Tassilo Barth, Michael Wiegand, Mittul Singh, and Dietrich Klakow. "Effective slot filling based on shallow distant supervision methods." *arXiv preprint arXiv:1401.1158* (2014).

[10] Muhič, Andrej, Jan Rupnik, and Primož Škraba. "Cross-lingual document similarity." In *Information Technology Interfaces (ITI), Proceedings of the ITI 2012 34th International Conference on*, pp. 387-392. IEEE, 2012.